DOCUMENT RESUME

ED 294 917 TM 011 649

AUTHOR Linacre, John Michael
TITLE The Practical Realization of the Standard Error of the Mantel-Haenszel Statistic.
PUB DATE Apr 88
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Error of Measurement; Estimation (Mathematics); *Item Analysis; Statistical Analysis; *Statistical Bias
IDENTIFIERS *Mantel Haenszel Procedure; Rasch PROX Procedure

ABSTRACT

Simulations were performed to verify the accuracy with which the Mantel-Haenszel (MH) and Rasch PROX procedures recover simulated item bias. Several standard error estimators for the MH procedure were evaluated. Item bias is recovered satisfactorily by both techniques under all simulated conditions. The proposed MH standard error estimators have different properties and are to be interpreted with caution. When using the MH procedure, the analyst is advised to calculate the size of the item bias and error variance using several sub-setting criteria and more than one standard error estimator. The PROX procedure provides a computationally independent and statistically conservative check on the estimates provided by the MH procedure. (Six tables show variations of item bias estimates across new score sub-sets.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294917

The practical realization of the standard error of the Mantel-Haenszel statistic

John Michael Linacre

University of Chicago

MESA Psychometric Laboratory

5835 S. Kimbark Ave.

Chicago Il 60637

Paper presented at the AERA Annual Meeting,

New Orleans April 6, 1988

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JOHN M. LINACRE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM011649

BEST COPY AVAILABLE

Abstract

Simulations were performed to verify the accuracy with which the Mantel-Haenszel (MH) and Rasch PROX procedures recover simulated item bias. Several standard error estimators for the MH procedure were evaluated. Item bias is recovered satisfactorily by both techniques under all simulated conditions.

The proposed MH standard error estimators have different properties and are to be interpreted with caution.

When using the MH procedure, the analyst is advised to calculate the size of the item bias and error variance using several subsetting criteria and more than one standard error estimator. The PROX procedure provides a computationally independent and statistically conservative check on the estimates provided by the MH procedure.

Key words: Item Bias, Mantel-Haenszel, Standard Error, PROX, Rasch

Introduction.

The identification of item bias is an important aspect of test analysis, and thus a vital component of testing policy and testing practice. The determination of which items in a test are biased, by how much, and against whom, can have a significant impact on how test results are perceived by the community, and how they are used as a selection device.

This paper compares standard error estimates proposed for the 1) Mantel-Haenszel (MH) statistic, 2) Rasch PROX statistic; two statistics which have been applied to the determination of item bias.

Through the use of simulated data, the accuracy of the bias statistics and their standard errors is obtained, and guidance given to the analyst as to the reliability of these statistics.

The Mantel-Haenszel (MH) Statistic

Mantel and Haenszel (1959) presented a chi-square test for factors possibly associated with the occurrence of a disease, which was based on 2x2 contingency tables. This MH procedure has been adopted by the educational community as one means of identifying biased test items.

In this paper, we are not concerned with the chi-square test of significance of bias, but rather with the associated log odds estimator of the size of the bias, θ (Phillips & Holland 1986). This is because, on the one hand, bias can be statistically significant but negligibly small. On the other hand, bias can be large, but fail to be statistically significant due to the small size of the population under examination.

We will be considering the situation in which there is one biased item in a relatively short test, given to a relatively small population. This simulates the conditions that may arise in tests relating to hiring or remedial programs.

The examinee population is perceived to consist of two groups: the Focus (minority) group against whom the suspect item is perceived to be biased, and the Reference (majority) group of the remainder of the population. They have the same ability distribution and the unbiased items of the test are uniformly distributed and targeted on the examinee population.

Creating population subsets with which to detect item bias

To calculate the MH θ statistic the reference (majority) group, R, and the focus (minority) group, F, are divided into

comparable subsets, in order to compare their performance on the suspected biased item.

These comparable subsets may be based on whatever information is available, but a common criterion is that comparable subsets obtained similar raw scores on the remainder of the test of which the biased item is a part.

As test length and size of the reference and focus groups may vary considerably between different testing situations, the manner in which raw scores are combined into subsets is based on the exigencies of the testing situation. The most precise grouping is obtained when each possible raw score corresponds to one sub-set. Then if there were a maximum score of 100 on a test, there would be 101 subsets of both the reference and focus groups. In each subset, a count is required of those who succeeded on the biased item and those who failed. Unless the population is large and diverse, there would be many zero counts. Consequently, it is often necessary to combine several adjacent scores into each subset.

For these simulations, the number of unbiased items administered was chosen such that the MH θ statistic could be based on one raw score level per subset, two adjacent score levels per subset, four adjacent score levels per subset, and so on until

the lower half of all possible raw scores were in one subset, and the upper half in the other.

Calculating the MH θ statistic.

Using the notation and exposition of Holland and Thayer (1986)', for each pair of the J comparable subsets of groups R and F:

		Performance on biased item		
		Succeeded	Failed	Total
Group	R (majority)	Aj	Bj	
	F (minority)	Cj	Dj	
	Total			Tj

Then the MH odds estimator, Φ, is

$$\Phi = R / S$$

where

$$R = \Sigma Rj = \Sigma AjDj/Tj \quad j=1,J$$

and $Rj = AjDj/Tj$

and

$$S = \Sigma Sj = \Sigma BjCj/Tj \quad j=1,J$$

and $Sj = BjCj/Tj$.

The log-odds estimator, θ, is

$$\theta = \log (R / S).$$

θ is measured in "logits" (log-odds units).

The Standard Error of the MH log-odds estimate.

There is only limited value in calculating the size of an estimate without determining how accurately it is known. The standard error (SE) of the MH log-odds estimate has not been unequivocally formulated. Rather several approximations have been presented.

The purpose of this paper is to use simulated data to determine the actual variance of θ in situations similar to those to be encountered in practice. These variances can then be compared to the calculated error variances, and conclusions drawn as to the utility of each of the SE equations.

The SE variance equations to be considered are those presented in Phillips and Holland (1986).

Hauck (1979):

$$VH = \frac{1}{S^2} \sum Sj^2 \left(\frac{1}{Aj} + \frac{1}{Bj} + \frac{1}{Cj} + \frac{1}{Dj} \right)$$

Breslow (1981):

$$VB = \frac{1}{R^2} \sum \frac{1}{Tj^2} (AjDj - \Phi BjCj)^2$$

Flanders (1985):

$$VF = \frac{1}{S^2} \sum \frac{BjCj}{Tj^2} \left(\frac{(Aj+Dj+1)}{\Phi} + (Bj+Cj-1) \right)$$

Phillips and Holland (1986):

$$VP = \frac{1}{2R^2} \sum \frac{1}{Tj^2} (AjDj + \Phi BjCj)(Aj + Dj + \Phi(Bj+Cj))$$

The Rasch PROX bias estimator.

The MH procedure can be shown to assume a probabilistic model of the same assumptions as the Rasch (one-parameter) IRT model (Linacre & Wright 1986). The advantage of the MH procedure over most IRT methods of log-odds estimation of item bias is that it is simple to calculate and non-iterative.

There is one IRT procedure which has properties almost identical to the MH procedure, the PROX estimation algorithm for the Rasch model. This algorithm determines ability estimates for each member of both the Focus and Reference groups, based on the unbiased items in the test. These ability estimates are then applied to the observed performance on the biased item to calculate its difficulty for the Focus group and its difficulty for the Reference group. The difference between these two estimates is the bias.

Estimates using this algorithm were also obtained for the simulated data.

The PROX Bias is calculated from

$$PB = DF - DR$$

where DF = difficulty of biased item for group F
and DR = difficulty of biased item for group R.

$$DF = MF + \sqrt{\left(1 + \frac{SF^2}{2.89}\right)} \log \left(\frac{\Sigma Dj}{\Sigma Cj} \right)$$

where MF = mean ability of group F

SF^2 = ability variance of group F

$$DR = MR + \sqrt{\left(1 + \frac{SR^2}{2.89}\right)} \log \left(\frac{\Sigma Bj}{\Sigma Aj} \right)$$

where MR = mean ability of group R

SR^2 = ability variance of group R

MF, SF, MR, SR may be calculated non-iteratively using the PROX procedures in Wright and Stone (1979 p. 34-44).

In these simulations, if an examinee (or item) obtained an extreme score (0 or maximum), the corresponding parameter would be inestimable. Consequently a Bayesian-type adjustment of one half score point was made to each of the relatively few extreme scores before PROX abilities or difficulties were calculated. This has the effect of raising the mean estimated ability of the examinees and attenuating their variance. This would make a biased item appear more biased.

The error variance of the PROX bias is:

$$PV = \left(1 + \frac{SF^2}{2.89}\right)\left(\frac{\Sigma(Cj+Dj)}{\Sigma\, Cj\ \Sigma\, Dj}\right) + \left(1 + \frac{SR^2}{2.89}\right)\left(\frac{\Sigma(Aj+Bj)}{\Sigma\, Aj\ \Sigma\, Bj}\right)$$

Simulating the data.

In order to test the accuracy of the MH log-odds estimator, θ, data was simulated according to a simple log-odds model. For each replication, two populations of examinees were simulated, the "majority" or reference group, R, and the "minority" or focus group, F. The ability parameter given for examinee n was Bn . A sample of "unbiased" test items was simulated. The difficulty parameter given item i was Di. The probability of success of examinee n on item i, Pni, was specified as:

$$\log(Pni/(1-Pni)) = Bn - Di.$$

For unbiased items, this formulation is the same whether the examinee is in the majority or minority groups.

Besides these unbiased items, in each simulation, there was inserted one biased item. For the majority group it had a difficulty parameter of Dmaj, and for the minority group a difficulty of Dmin. Thus

$$Bias = Dmin - Dmaj$$

This bias was generally specified to be positive (biased against the minority group).

The probability of success of examinee n of the majority group on the biased item was determined by

$$\log(Pni/(1-Pni)) = Bn - Dmaj$$

and the probability of success of examinee n' of the minority group on the biased item was determine by

$$\log(Pn'i/(1-Pn'i)) = Bn' - Dmin$$

The simulated data were generated by comparing the probabilities obtained using the log-odds formulation with uniform random numbers in the interval (0,1). If the probability were greater than the random number, the simulated data point was recorded as a success, otherwise a failure.

The distribution of person ability parameters was normal random with mean 0 and standard deviation 1. The distribution of the difficulty parameters of the unbiased items was uniform over the interval -1 to +1. For the reference group, the biased item

was targeted on the mean of the person distribution (Dmaj = 0). Its difficulty for the Focus group was simulated at various levels between -0.5 (favorable) and 3.0 (very unfavorable).

For each simulation, 100 replications of the person and item parameters, and simulated examinee responses were performed. There were a very few replications in which no one in the Focus group succeeded on the biased item. Such replications were ignored, and the simulation continued until 100 successful replications were obtained. Omitting such extreme replications, (which are inestimable using the MH technique), would tend to attenuate the degree of bias, and its variance, reported by the simulation technique.

Recovery of the generators:

The simulations were designed to answer several questions about the degree to which the MH log odds estimator recovered the generating item bias parameter, and the extent to which the several error variance algorithms reflect the observed variance in the bias estimates.

I. How accurately do the procedures recover the generating item bias ?

a) According to different population sizes

b) According to different test lengths

c) According to different degrees of bias

d) According to different sub-sets of score levels.

II. Which of the proposed error variance algorithms correspond most closely with the actual bias variance ?

Recovering the simulated item bias

Tables 1 to 3 show the results of the simulations for different combinations of these four factors. In Table 1, the Focus and Reference group are each same size (100 examinees). In Table 2, the Focus group comprises one quarter of the population (50 in focus group, 150 in Reference group). In Table 3, the Focus group comprises one eighth of the population (25 in Focus group, 175 in Reference group). This wide range of relative group sizes reflects the conditions under which the MH procedure would be expected to function in practice.

Two different test lengths were simulated: a test of 32 items (31 unbiased and 1 biased) and 64 items (63 unbiased and 1 biased). We will discuss later what would occur if even longer tests were simulated.

Bias was simulated from 0.0 (unbiased) to 3.0 (very biased against Focus group), and also -0.5 (slightly biased in favor of Focus group), where meaningful. Items as biased as much as 2.0 or 3.0 logits are outside the range of the unbiased test and reflect very aberrant situations. 0.5 and 1.0 logits of bias more closely reflect the degree of bias which the investigator would wish to identify and quantify.

The number of unbiased test items was selected so as to obtain a broad range of subsetting choices. 31 unbiased items represent 32 score levels (0 to 31). For creating subsets these were divided as follows: 1 score level per subset (each subset is 1/32 of possible score levels); 2 score levels per subset (1/16 of score levels); 4 score level per subset (1/8 of score levels); and so on up to 16 score levels per subset (1/2 of score levels per subset per group, one subset scored in the range 0-15, the other subset in the range 16-31). The test of 63 unbiased items was divided into subsets in a similar way form 1/64 to 1/2. This gives a range of subsetting from the finest possible, with the information available, to the coarsest.

The PROX method is also included for comparison. It considers each group as a whole, and so subsetting is unnecessary.

The first analysis in Table 1 corresponds to an unbiased item considered for Reference and Focus groups which are of the same size for the short (31 unbiased items) test. Below it, the same conditions are shown for the long (63 unbiased items) test.

In these tables, "Bias mean" refers to the mean of the log-odds bias estimate for the 100 replications. "Bias S.D." is the standard deviation of the bias estimates over the 100 replications, and so is the empirically-determined standard error of the Bias mean. "Stand. Diff." is the standardized difference between the observed mean bias and the simulated bias, obtained by dividing difference between the observed and simulated biases by the bias standard error. Any amount of absolute value less than 1 indicates that the mean observed bias is less than one standard error away from the simulated bias.

Subsetting has been reported in terms of the fraction of score levels included in each subset, so that equivalent sized subsets for simulations with different test lengths are aligned vertically.

Inspection of these tables reveal no standardized difference of size greater than 1, and both the MH procedure and PROX have recovered the simulated item bias within reasonable expectations. Nevertheless, general trends are discernible. As the degree of bias becomes greater, the accuracy of recovery of the bias deteriorates, but whether the bias is over- or under-estimated depends on the local conditions.

Longer test length (number of unbiased items) does not automatically mean greater accuracy in estimating item bias. The MH procedure is based on 2x2 contingency tables for each subset. If the number of items is large there will be many empty cells in the tables unless the subsetting criteria are coarse. The subsetting has been reported by fraction of score levels so that the effect of increasing test length and making finer subsetting (smaller numbers of examinees per subset) can be observed. The results indicate that increasing test length does not automatically increase the accuracy of the bias estimate.

As predicted by Linacre and Wright (1986), the manner in which subsetting is performed does effect the size of the estimate. This can be seen in extreme form by inspecting the end of Table 3. Depending on the manner in which subsets are formed, the bias may be over or underestimated. Over the simulations as a whole, there does not appear to be a method of subsetting which is appropriate, a priori, to all conditions. Consequently the results obtained are dependent, to some extent, on the preferences of the analyst.

As log odds estimators, there is no clear preference between PROX and MH. Under nearly all conditions, the PROX estimate is not as bad as the worst MH estimate, and, under some conditions, it is better than the best MH estimate. Only when the bias is so great that it takes the item outside the range of the unbiased does the PROX estimate become worse than any MH estimate.

Accuracy of error variance estimates.

Having confirmed that the generating bias is successfully recovered, let us determine how closely the proposed error variance calculations agree with the variance of the estimated bias as empirically observed over 100 replications.

Tables 4, 5 and 6 show the error variance estimates corresponding to the log odds bias estimates just discussed. The tables are constructed in the same way as before so far as the column headings and "Bias mean" are concerned. "Bias variance" is the variance of the bias estimates over the 100 replications and is the empirical realization of the error variance. "VB" is the mean of Breslow's error variance estimate across the 100 replications. "VF" is the mean of Flanders' estimate. "VH", where shown, is Hauck's estimate, and "VP", where shown, is the mean of Phillips and Holland's estimate. "Variance/VB" is the ratio of the observed variance to the VB variance, similarly for "Variance/VF". The final column gives the equivalent information for the PROX estimates.

VH is only shown for the first two analyses. It is generally inestimable because it relies on every cell in each subset's 2x2 contingency table being non-zero. This only occurs when the each subset contains a sufficiently large range of score levels. The

VH estimates shown are representative of, or better than, the VH results obtained from the other replications. VH is clearly far from satisfactory.

VP is only listed when it differs from VF for any subsetting by more than .005. In general this occurs only when the item bias is large. In those cases when VF and VP differ noticeably, there is no clear pattern of superiority of one over the other, except that in really extreme cases (such as at the end of Table 6), VP overestimates the variance considerably more than VF. Since VP is a later and computationally slightly more complex estimator, VF will be preferred in this paper.

VB is earlier and computationally slightly simpler then VF. VB estimates a systematically smaller error variance than VF. However, as inspection of the variance ratios indicate, this does not mean that one or other estimator is universally preferable. The VF estimate, however, is more consistent across different subsets. In general, VF is preferable to VB. Depending on the simulation parameters, VF may be above or below the observed variance, but is generally close to it.

The PROX error variance is also generally close to the observed variance, but usually overestimates the standard error.

Conclusion

These simulations have confirmed the utility of the Mantel-Haenszel procedure in estimating item bias, under those circumstances where the constraints of a log-odds model hold.

However, these simulations also indicate that the quality of both bias and standard error estimates are dependent on the nature of the local situation and, with respect to subsetting, on the arbitrary choices of the analyst.

Hauck's error variance estimator is not trustworthy. Flanders' estimator is better than Phillips and Holland's because of its better behavior under extreme conditions.

The conservative analytic approach to using the MH procedure is to calculate the estimated item bias using a range of subsetting choices and with both Breslow's and Flanders' standard error variance estimators. Comparison of the pattern of results obtained with those given in this paper may serve as a guide to the true magnitude of the underlying item bias.

The PROX algorithm provides a computationally independent estimate of the same bias and is to be recommended where results of the item analysis are in dispute.

Bibliography

Holland, P.W., Thayer, D.T. (1986) Differential Item functioning and the Mantel-Haenszel Procedure. Princeton, NJ: ETS Technical Report No. 86-69

Linacre, J.M., Wright, B.D. (1986) Item bias: Mantel-Haenszel and the Rasch Model. Chicago: University of Chicago, MESA Psychometric Laboratory, Memorandum no. 39

Mantel, N., Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748

Phillips, A., Holland, P.W. (1986) A new estimator for the variance of the Mantel-Haenszel Log-Odds-Ratio Estimator. Princeton, NJ: ETS Technical Report No. 86-67

Wright, B.D., Stone, M.H.. (1979) Best Test Design. Chicago, Il: MESA Press

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 100, Reference Group = 100									
Bias mean	31	0.00		-0.002	0.002	0.002	-0.007	-0.031	-0.015
Bias S.E.				0.300	0.297	0.292	0.288	0.282	0.298
Stand. Diff.				-0.01	0.01	0.01	-0.03	-0.11	-0.05
Bias mean	63	0.00	-0.020	-0.008	-0.005	0.007	0.002	-0.002	0.011
Bias S.E.			0.348	0.330	0.319	0.296	0.285	0.278	0.283
Stand. Diff.			-0.06	-0.02	-0.02	0.02	0.01	-0.01	0.04
Bias mean	31	0.50		0.494	0.491	0.490	0.470	0.436	0.488
Bias S.E.				0.310	0.314	0.302	0.306	0.290	0.301
Stand. Diff.				-0.02	-0.03	-0.03	-0.10	-0.22	-0.04
Bias mean	63	0.50	0.544	0.541	0.531	0.542	0.537	0.526	0.524
Bias S.E.			0.406	0.371	0.349	0.336	0.334	0.334	0.329
Stand. Diff.			0.11	0.11	0.09	0.13	0.11	0.08	0.07
Bias mean	31	1.00		0.946	0.935	0.929	0.922	0.898	1.033
Bias S.E.				0.317	0.316	0.307	0.311	0.300	0.332
Stand. Diff.				-0.17	-0.20	-0.23	-0.25	-0.34	0.10
Bias mean	63	1.00	0.979	0.982	0.964	0.962	0.943	0.912	0.958
Bias S.E.			0.378	0.342	0.329	0.312	0.308	0.300	0.296
Stand. Diff.			-0.06	-0.05	-0.11	-0.12	-0.19	-0.29	-0.14
Outside range of unbiased test:									
Bias mean	31	2.00		2.149	2.125	2.098	2.069	1.973	2.104
Bias S.E.				0.463	0.427	0.404	0.386	0.346	0.362
Stand. Diff.				0.32	0.29	0.24	0.18	-0.08	0.29
Bias mean	63	2.00	1.975	1.934	1.931	1.939	1.920	1.840	1.965
Bias S.E.			0.482	0.439	0.402	0.389	0.387	0.366	0.360
Stand. Diff.			-0.05	-0.15	-0.17	-0.16	-0.21	-0.44	-0.10
Bias mean	31	3.00		2.906	2.911	2.882	2.845	2.746	3.071
Bias S.E.				0.505	0.454	0.441	0.417	0.405	0.449
Stand. Diff.				-0.19	-0.20	-0.27	-0.37	-0.63	0.16
Bias mean	63	3.00	3.002	3.068	3.057	3.036	3.014	2.905	3.154
Bias S.E.			0.696	0.643	0.604	0.620	0.603	0.586	0.640
Stand. Diff.			0.00	0.11	0.09	0.06	0.02	-0.16	0.24

Table 1: Variation of item bias estimate across raw score subsets.
Simulations based on Focus group size = 100, Reference group = 100,
both groups with normal random distribution (0,1).
Unbiased items are uniformly distributed (-1,1).
Biased item with difficulty 0 (Reference), simulated bias (Focus).
Bias estimates as observed from 100 replications

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 50, Reference Group = 150									
Bias mean	31	-0.50		-0.468	-0.453	-0.456	-0.441	-0.403	-0.441
Bias S.E.				0.412	0.397	0.384	0.375	0.340	0.402
Stand. Diff.				0.08	0.12	0.12	0.16	0.29	0.15
Bias mean	63	-0.50	-0.597	-0.582	-0.571	-0.584	-0.572	-0.558	-0.566
Bias S.E.			0.448	0.422	0.411	0.395	0.384	0.374	0.379
Stand. Diff.			-0.22	-0.19	-0.17	-0.21	-0.19	-0.15	-0.17
Bias mean	31	0.00		-0.004	0.018	0.013	-0.002	-0.030	0.001
Bias S.E.				0.389	0.347	0.331	0.326	0.305	0.335
Stand. Diff.				-0.01	0.05	0.04	-0.00	-0.10	0.00
Bias mean	63	0.00	-0.025	-0.026	-0.020	-0.014	-0.022	-0.033	-0.007
Bias S.E.			0.404	0.405	0.401	0.389	0.379	0.367	0.363
Stand. Diff.			-0.06	-0.06	-0.05	-0.04	-0.06	-0.09	-0.02
Bias mean	31	0.50		0.494	0.496	0.486	0.481	0.462	0.507
Bias S.E.				0.383	0.367	0.358	0.361	0.347	0.369
Stand. Diff.				-0.01	-0.01	-0.04	-0.05	-0.11	0.02
Bias mean	63	0.50	0.458	0.483	0.491	0.471	0.451	0.424	0.474
Bias S.E.			0.454	0.429	0.421	0.402	0.386	0.371	0.401
Stand. Diff.			-0.09	-0.04	-0.02	-0.07	-0.13	-0.20	-0.06
Bias mean	31	1.00		0.942	0.941	0.937	0.921	0.876	0.934
Bias S.E.				0.414	0.407	0.382	0.376	0.356	0.401
Stand. Diff.				-0.14	-0.15	-0.16	-0.21	-0.35	-0.16
Bias mean	63	1.00	1.041	1.022	1.019	1.031	1.030	0.990	1.025
Bias S.E.			0.495	0.435	0.431	0.433	0.424	0.396	0.424
Stand. Diff.			0.08	0.05	0.04	0.07	0.07	-0.03	0.06
Outside range of unbiased test:									
Bias mean	31	2.00		2.136	2.124	2.096	2.052	1.973	2.152
Bias S.E.				0.586	0.560	0.531	0.499	0.466	0.563
Stand. Diff.				0.23	0.22	0.18	0.10	-0.06	0.27
Bias mean	63	2.00	2.081	2.073	2.044	2.054	2.026	1.924	2.027
Bias S.E.			0.651	0.562	0.506	0.505	0.498	0.456	0.453
Stand. Diff.			0.13	0.13	0.09	0.11	0.05	-0.17	0.06
Bias mean	31	3.00		3.033	3.040	3.003	2.983	2.846	3.147
Bias S.E.				0.786	0.793	0.753	0.772	0.657	0.727
Stand. Diff.				0.04	0.05	0.00	-0.02	-0.23	0.20
Bias mean	63	3.00	3.167	3.210	3.220	3.203	3.161	3.015	3.271
Bias S.E.			0.715	0.716	0.724	0.761	0.722	0.642	0.693
Stand. Diff.			0.23	0.29	0.30	0.27	0.22	0.02	0.39

Table 2: Variation of item bias estimate across raw score subsets.
Simulations based on Focus group size = 50, Reference group = 150,
both groups with normal random distribution (0,1).
Unbiased items are uniformly distributed (-1,1).
Biased item with difficulty 0 (Reference), simulated bias (Focus).
Bias estimates as observed from 100 replications

Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 25, Reference Group = 175								
Bias mean	31	-0.50	-0.468	-0.475	-0.487	-0.467	-0.407	-0.521
Bias S.E.			0.577	0.559	0.529	0.505	0.476	0.584
Stand. Diff.			0.05	0.04	0.03	0.06	0.20	-0.04
Bias mean	63	-0.50	-0.437	-0.462	-0.483	-0.470	-0.475	-0.457	-0.469
Bias S.E.			0.633	0.503	0.471	0.452	0.445	0.418	0.431
Stand. Diff.			0.10	0.08	0.04	0.07	0.06	0.10	0.07
Bias mean	31	0.00	0.042	0.027	0.027	0.042	0.045	0.022
Bias S.E.			0.472	0.464	0.466	0.448	0.428	0.460
Stand. Diff.			0.09	0.06	0.06	0.09	0.10	0.05
Bias mean	63	0.00	0.025	0.021	-0.019	-0.026	-0.031	-0.051	-0.017
Bias S.E.			0.587	0.524	0.514	0.496	0.474	0.451	0.457
Stand. Diff.			0.04	0.04	-0.04	-0.05	-0.06	-0.11	-0.04
Bias mean	31	0.50	0.514	0.514	0.506	0.500	0.462	0.479
Bias S.E.			0.541	0.523	0.475	0.476	0.460	0.490
Stand. Diff.			0.03	0.03	0.01	-0.00	-0.08	-0.04
Bias mean	63	0.50	0.628	0.624	0.610	0.599	0.591	0.578	0.562
Bias S.E.			0.548	0.513	0.503	0.479	0.486	0.448	0.476
Stand. Diff.			0.23	0.24	0.22	0.21	0.19	0.17	0.13
Bias mean	31	1.00	1.075	1.051	1.051	1.033	1.007	1.057
Bias S.E.			0.702	0.653	0.619	0.587	0.565	0.630
Stand. Diff.			0.11	0.08	0.08	0.06	0.01	0.09
Bias mean	63	1.00	0.952	0.935	0.934	0.940	0.904	0.868	0.917
Bias S.E.			0.619	0.505	0.464	0.466	0.456	0.462	0.465
Stand. Diff.			-0.08	-0.13	-0.14	-0.13	-0.21	-0.29	-0.18
Outside range of unbiased test:								
Bias mean	31	2.00	2.001	2.006	2.038	1.997	1.899	2.079
Bias S.E.			0.676	0.649	0.673	0.665	0.615	0.681
Stand. Diff.			0.00	0.01	0.06	-0.00	-0.16	0.12
Bias mean	63	2.00	2.117	2.151	2.206	2.138	2.068	2.013	2.132
Bias S.E.			0.826	0.793	0.846	0.786	0.725	0.725	0.719
Stand. Diff.			0.14	0.19	0.24	0.18	0.09	0.02	0.18
Bias mean	31	3.00	2.861	2.974	2.957	3.012	2.839	3.072
Bias S.E.			0.666	0.755	0.774	0.817	0.659	0.696
Stand. Diff.			-0.21	-0.03	-0.06	0.01	-0.24	0.10
Bias mean	63	3.00	2.650	2.767	2.933	2.982	3.036	2.829	3.016
Bias S.E.			0.655	0.711	0.801	0.789	0.853	0.689	0.716
Stand. Diff.			-0.54	-0.33	-0.08	-0.02	0.04	-0.25	0.02

Table 3: Variation of item bias estimate across raw score subsets.
Simulations based on Focus group size = 25, Reference group = 175,
both groups with normal random distribution (0,1).
Unbiased items are uniformly distributed (-1,1).
Biased item with difficulty 0 (Reference), simulated bias (Focus).
Bias estimates as observed from 100 replications

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 100, Reference Group = 100									
Bias mean	31	0.00		-0.002	0.002	0.002	-0.007	-0.031	-0.015
Bias variance				0.090	0.088	0.085	0.083	0.080	0.089
VB or PROX SE				0.103	0.091	0.077	0.062	0.044	0.116
Variance/VB or PROX SE				0.88	0.96	1.12	1.35	1.80	0.77
VF				0.110	0.102	0.097	0.094	0.089	
Variance/VF				0.82	0.87	0.88	0.88	0.90	
VH				n.e.	n.e.	n.e.	2.733	0.364	
Bias mean	63	0.00	-0.020	-0.008	-0.005	0.007	0.002	-0.002	0.011
Bias variance			0.121	0.109	0.102	0.088	0.081	0.078	0.080
VB or PROX SE			0.125	0.104	0.092	0.083	0.065	0.048	0.112
Variance/VB or PROX SE			0.97	1.05	1.11	1.06	1.25	1.60	0.72
VF			0.132	0.112	0.104	0.100	0.096	0.092	
Variance/VF			0.92	0.97	0.98	0.88	0.84	0.84	
VH			n.e.	n.e.	n.e.	n.e.	2.995	0.377	
Bias mean	31	0.50		0.494	0.491	0.490	0.470	0.436	0.488
Bias variance				0.096	0.098	0.091	0.093	0.084	0.091
VB or PROX SE				0.109	0.099	0.091	0.079	0.060	0.119
Variance/VB or PROX SE				0.89	1.00	1.00	1.18	1.40	0.76
VF				0.114	0.105	0.101	0.097	0.092	
Variance/VF				0.84	0.94	0.91	0.96	0.91	
Bias mean	63	0.50	0.544	0.541	0.531	0.542	0.537	0.526	0.524
Bias variance			0.165	0.138	0.122	0.113	0.112	0.111	0.108
VB or PROX SE			0.130	0.105	0.093	0.082	0.053	0.037	0.114
Variance/VB or PROX SE			1.27	1.31	1.31	1.38	2.09	2.97	0.95
VF			0.137	0.115	0.106	0.102	0.098	0.094	
Variance/VF			1.21	1.20	1.15	1.11	1.14	1.18	
Bias mean	31	1.00		0.946	0.935	0.929	0.922	0.898	1.033
Bias variance				0.101	0.100	0.094	0.097	0.090	0.110
VB or PROX SE				0.116	0.103	0.099	0.067	0.040	0.129
Variance/VB or PROX SE				0.87	0.97	0.96	1.45	2.27	0.85
VF				0.121	0.110	0.105	0.102	0.098	
Variance/VF				0.83	0.91	0.90	0.95	0.92	
Bias mean	63	1.00	0.979	0.982	0.964	0.962	0.943	0.912	0.958
Bias variance			0.143	0.117	0.108	0.097	0.095	0.090	0.087
VB or PROX SE			0.143	0.119	0.103	0.091	0.068	0.044	0.122
Variance/VB or PROX SE			1.00	0.98	1.05	1.06	1.40	2.04	0.72
VF			0.146	0.124	0.113	0.109	0.105	0.100	
Variance/VF			0.98	0.94	0.95	0.90	0.90	0.90	

Table 4. Continues.

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 100, Reference Group = 100									
Outside range of unbiased test:									
Bias mean	31	2.00		2.149	2.125	2.098	2.069	1.973	2.104
Bias variance				0.214	0.182	0.163	0.149	0.119	0.131
VB or PROX SE				0.188	0.148	0.130	0.104	0.077	0.174
Variance/VB or PROX SE				1.13	1.23	1.26	1.43	1.56	0.75
VF				0.201	0.172	0.160	0.152	0.138	
Variance/VF				1.06	1.05	1.02	0.98	0.86	
VP				0.207	0.178	0.164	0.155	0.141	
Bias mean	63	2.00	1.975	1.934	1.931	1.939	1.920	1.840	1.965
Bias variance			0.232	0.193	0.162	0.151	0.149	0.134	0.129
VB or PROX SE			0.237	0.179	0.144	0.127	0.102	0.074	0.165
Variance/VB or PROX SE			0.98	1.08	1.12	1.19	1.46	1.81	0.78
VF			0.241	0.187	0.164	0.155	0.149	0.136	
Variance/VF			0.96	1.03	0.99	0.98	1.01	0.99	
Bias mean	31	3.00		2.906	2.911	2.882	2.845	2.746	3.071
Bias variance				0.255	0.206	0.194	0.174	0.164	0.201
VB or PROX SE				0.322	0.247	0.205	0.173	0.109	0.287
Variance/VB or PROX SE				0.79	0.83	0.95	1.00	1.50	0.70
VF				0.344	0.282	0.256	0.243	0.223	
Variance/VF				0.74	0.73	0.76	0.71	0.74	
VP				0.366	0.299	0.269	0.252	0.230	
Bias mean	63	3.00	3.002	3.068	3.057	3.036	3.014	2.905	3.154
Bias variance			0.484	0.414	0.365	0.384	0.363	0.343	0.410
VB or PROX SE			0.467	0.378	0.301	0.254	0.197	0.115	0.327
Variance/VB or PROX SE			1.04	1.10	1.21	1.52	1.85	2.98	1.25
VF			0.483	0.399	0.340	0.310	0.290	0.258	
Variance/VF			1.00	1.04	1.07	1.24	1.25	1.33	
VP			0.515	0.427	0.359	0.327	0.307	0.272	

Table 4: Behavior of standard error variance estimates across subsets.
Simulations based on Focus group size = 100, Reference group = 100,
both groups with normal random distribution (0,1).
Unbiased items are uniformly distributed (-1,1).
Biased item with difficulty 0 (Reference), simulated bias (Focus).
Bias estimates as observed from 100 replications

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 50, Reference Group = 150									
Bias mean	31	-0.50		-0.468	-0.453	-0.456	-0.441	-0.403	-0.441
Bias variance				0.169	0.158	0.147	0.141	0.115	0.162
VB or PROX SE				0.130	0.109	0.100	0.086	0.070	0.161
Variance/VB or PROX SE				1.31	1.45	1.47	1.64	1.65	1.00
VF				0.150	0.138	0.133	0.128	0.119	
Variance/VF				1.13	1.14	1.11	1.10	0.97	
Bias mean	63	-0.50	-0.597	-0.582	-0.571	-0.584	-0.572	-0.558	-0.566
Bias variance			0.200	0.178	0.169	0.156	0.147	0.140	0.144
VB or PROX SE			0.175	0.147	0.128	0.116	0.083	0.062	0.156
Variance/VB or PROX SE			1.15	1.21	1.32	1.35	1.77	2.27	0.92
VF			0.190	0.162	0.149	0.142	0.136	0.130	
Variance/VF			1.05	1.10	1.14	1.10	1.08	1.08	
Bias mean	31	0.00		-0.004	0.018	0.013	-0.002	-0.030	0.001
Bias variance				0.151	0.121	0.110	0.106	0.093	0.112
VB or PROX SE				0.142	0.119	0.103	0.069	0.044	0.156
Variance/VB or PROX SE				1.06	1.02	1.06	1.54	2.12	0.72
VF				0.148	0.136	0.130	0.125	0.119	
Variance/VF				1.02	0.89	0.84	0.85	0.78	
Bias mean	63	0.00	-0.025	-0.026	-0.020	-0.014	-0.022	-0.033	-0.007
Bias variance			0.163	0.164	0.161	0.151	0.144	0.135	0.132
VB or PROX SE			0.170	0.151	0.140	0.122	0.088	0.068	0.150
Variance/VB or PROX SE			0.96	1.09	1.15	1.24	1.63	1.98	0.87
VF			0.175	0.148	0.138	0.133	0.129	0.123	
Variance/VF			0.93	1.11	1.16	1.14	1.11	1.09	
Bias mean	31	0.50		0.494	0.496	0.486	0.481	0.462	0.507
Bias variance				0.147	0.134	0.128	0.131	0.120	0.136
VB or PROX SE				0.145	0.129	0.115	0.073	0.051	0.162
Variance/VB or PROX SE				1.01	1.05	1.12	1.79	2.36	0.84
VF				0.153	0.142	0.135	0.132	0.126	
Variance/VF				0.96	0.95	0.94	0.99	0.95	
Bias mean	63	0.50	0.458	0.483	0.491	0.471	0.451	0.424	0.474
Bias variance			0.207	0.184	0.177	0.162	0.149	0.138	0.161
VB or PROX SE			0.178	0.152	0.128	0.119	0.095	0.056	0.157
Variance/VB or PROX SE			1.16	1.21	1.39	1.36	1.56	2.44	1.03
VF			0.180	0.154	0.143	0.137	0.133	0.127	
Variance/VF			1.15	1.19	1.24	1.18	1.13	1.09	
Bias mean	31	1.00		0.942	0.941	0.937	0.921	0.876	0.934
Bias variance				0.171	0.166	0.146	0.141	0.127	0.161
VB or PROX SE				0.155	0.136	0.117	0.090	0.062	0.176
Variance/VB or PROX SE				1.10	1.22	1.25	1.57	2.04	0.91
VF				0.171	0.157	0.150	0.144	0.136	
Variance/VF				1.00	1.06	0.98	0.98	0.93	

Table 5. Continues.

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 50, Reference Group = 150									
Bias mean	63	1.00	1.041	1.022	1.019	1.031	1.030	0.990	1.025
Bias variance			0.245	0.189	0.186	0.188	0.180	0.157	0.180
VB or PROX SE			0.200	0.164	0.142	0.136	0.100	0.083	0.175
Variance/VB or PROX SE			1.22	1.15	1.30	1.38	1.81	1.90	1.03
VF			0.208	0.175	0.161	0.156	0.151	0.144	
Variance/VF			1.18	1.08	1.15	1.20	1.19	1.09	
Outside range of unbiased test:									
Bias mean	31	2.00		2.136	2.124	2.096	2.052	1.973	2.152
Bias variance				0.344	0.314	0.282	0.249	0.217	0.316
VB or PROX SE				0.299	0.259	0.217	0.161	0.119	0.297
Variance/VB or PROX SE				1.15	1.21	1.30	1.55	1.83	1.07
VF				0.313	0.280	0.259	0.245	0.225	
Variance/VF				1.10	1.12	1.09	1.01	0.97	
VP				0.334	0.295	0.269	0.251	0.230	
Bias mean	63	2.00	2.081	2.073	2.044	2.054	2.026	1.924	2.027
Bias variance			0.423	0.316	0.256	0.255	0.248	0.208	0.205
VB or PROX SE			0.363	0.274	0.227	0.198	0.159	0.108	0.261
Variance/VB or PROX SE			1.17	1.15	1.13	1.29	1.56	1.92	0.79
VF			0.379	0.299	0.261	0.250	0.239	0.211	
Variance/VF			1.12	1.05	0.98	1.02	1.04	0.99	
VP			0.402	0.314	0.273	0.261	0.249	0.219	
Bias mean	31	3.00		3.033	3.040	3.003	2.983	2.846	3.147
Bias variance				0.618	0.629	0.567	0.596	0.432	0.529
VB or PROX SE				0.549	0.469	0.410	0.311	0.214	0.583
Variance/VB or PROX SE				1.12	1.34	1.38	1.92	2.02	0.91
VF				0.584	0.522	0.486	0.464	0.419	
Variance/VF				1.06	1.20	1.17	1.28	1.03	
VP				0.698	0.622	0.556	0.558	0.456	
Bias mean	63	3.00	3.167	3.210	3.220	3.203	3.161	3.015	3.271
Bias variance			0.511	0.513	0.524	0.579	0.521	0.412	0.480
VB or PROX SE			0.723	0.571	0.479	0.415	0.358	0.224	0.608
Variance/VB or PROX SE			0.71	0.90	1.09	1.40	1.46	1.84	0.79
VF			0.747	0.611	0.545	0.517	0.502	0.458	
Variance/VF			0.68	0.84	0.96	1.12	1.04	0.90	
VP			0.858	0.724	0.665	0.641	0.587	0.496	

Table 5: Behavior of standard error variance estimates across subsets.
Simulations based on Focus group size = 50, Reference group = 150,
both groups with normal random distribution (0,1).
Unbiased items are uniformly distributed (-1,1).
Biased item with difficulty 0 (Reference), simulated bias (Focus).
Bias estimates as observed from 100 replications

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 25, Reference Group = 175									
Bias mean	31	-0.50		-0.468	-0.475	-0.487	-0.467	-0.407	-0.521
Bias variance				0.333	0.312	0.280	0.255	0.226	0.341
VB or PROX SE				0.224	0.199	0.179	0.139	0.100	0.290
Variance/VB or PROX SE				1.48	1.57	1.56	1.83	2.26	1.18
VF				0.257	0.239	0.228	0.219	0.204	
Variance/VF				1.29	1.30	1.22	1.16	1.11	
Bias mean	63	-0.50	-0.437	-0.462	-0.483	-0.470	-0.475	-0.457	-0.469
Bias variance			0.400	0.253	0.222	0.205	0.198	0.175	0.186
VB or PROX SE			0.302	0.250	0.220	0.211	0.163	0.131	0.270
Variance/VB or PROX SE			1.32	1.01	1.01	0.97	1.21	1.33	0.69
VF			0.351	0.284	0.260	0.246	0.240	0.229	
Variance/VF			1.14	0.89	0.85	0.83	0.82	0.76	
VP			0.340	0.279	0.256	0.244	0.237	0.227	
Bias mean	31	0.00		0.042	0.027	0.027	0.042	0.045	0.022
Bias variance				0.222	0.215	0.217	0.200	0.183	0.212
VB or PROX SE				0.248	0.224	0.192	0.161	0.125	0.273
Variance/VB or PROX SE				0.90	0.96	1.13	1.25	1.46	0.78
VF				0.260	0.240	0.231	0.222	0.205	
Variance/VF				0.85	0.89	0.94	0.90	0.89	
Bias mean	63	0.00	0.025	0.021	-0.019	-0.026	-0.031	-0.051	-0.017
Bias variance			0.344	0.274	0.264	0.246	0.225	0.203	0.208
VB or PROX SE			0.304	0.248	0.217	0.188	0.123	0.096	0.259
Variance/VB or PROX SE			1.13	1.11	1.22	1.31	1.83	2.12	0.80
VF			0.324	0.271	0.253	0.240	0.231	0.221	
Variance/VF			1.06	1.01	1.04	1.02	0.97	0.92	
Bias mean	31	0.50		0.514	0.514	0.506	0.500	0.462	0.479
Bias variance				0.293	0.273	0.226	0.227	0.211	0.240
VB or PROX SE				0.246	0.208	0.185	0.162	0.134	0.282
Variance/VB or PROX SE				1.19	1.31	1.22	1.40	1.58	0.85
VF				0.268	0.248	0.235	0.229	0.217	
Variance/VF				1.09	1.10	0.96	0.99	0.97	
Bias mean	63	0.50	0.628	0.624	0.610	0.599	0.591	0.578	0.562
Bias variance			0.301	0.263	0.253	0.229	0.236	0.201	0.227
VB or PROX SE			0.313	0.266	0.242	0.200	0.147	0.116	0.276
Variance/VB or PROX SE			0.96	0.99	1.05	1.15	1.61	1.73	0.82
VF			0.323	0.274	0.256	0.245	0.237	0.223	
Variance/VF			0.93	0.96	0.99	0.93	1.00	0.90	

Table 6. Continues.

	Unbiased Items	Bias Simulated	Mantel-Haenszel Bias Estimation Fraction of score levels per subset 1/64	1/32	1/16	1/8	1/4	1/2	PROX Bias Est.
Population size: Focus Group = 25, Reference Group = 175									
Bias mean	31	1.00		1.075	1.051	1.051	1.033	1.007	1.057
Bias variance				0.493	0.426	0.383	0.345	0.319	0.397
VB or PROX SE				0.302	0.265	0.240	0.178	0.125	0.331
Variance/VB or PROX SE				1.63	1.61	1.59	1.94	2.55	1.20
VF				0.325	0.297	0.284	0.268	0.250	
Variance/VF				1.52	1.43	1.35	1.29	1.28	
VP				0.336	0.302	0.286	0.272	0.254	
Bias mean	63	1.00	0.952	0.935	0.934	0.940	0.904	0.868	0.917
Bias variance			0.383	0.255	0.215	0.217	0.208	0.213	0.216
VB or PROX SE			0.349	0.282	0.236	0.212	0.164	0.123	0.299
Variance/VB or PROX SE			1.10	0.91	0.91	1.02	1.27	1.73	0.72
VF			0.366	0.304	0.277	0.266	0.255	0.243	
Variance/VF			1.05	0.84	0.77	0.81	0.81	0.88	
VP			0.381	0.312	0.284	0.272	0.259	0.246	
Outside range of unbiased test:									
Bias mean	31	2.00		2.001	2.006	2.038	1.997	1.899	2.079
Bias variance				0.457	0.422	0.453	0.443	0.378	0.464
VB or PROX SE				0.487	0.407	0.339	0.256	0.177	0.538
Variance/VB or PROX SE				0.94	1.04	1.34	1.73	2.13	0.86
VF				0.508	0.461	0.448	0.431	0.407	
Variance/VF				0.90	0.92	1.01	1.03	0.93	
VP				0.548	0.497	0.492	0.467	0.428	
Bias mean	63	2.00	2.117	2.151	2.206	2.138	2.068	2.013	2.132
Bias variance			0.682	0.629	0.716	0.617	0.526	0.526	0.516
VB or PROX SE			0.622	0.531	0.483	0.406	0.316	0.174	0.555
Variance/VB or PROX SE			1.10	1.18	1.48	1.52	1.66	3.02	0.93
VF			0.644	0.566	0.537	0.479	0.457	0.426	
Variance/VF			1.06	1.11	1.33	1.29	1.15	1.24	
VP			0.725	0.633	0.614	0.540	0.493	0.457	
Bias mean	31	3.00		2.861	2.974	2.957	3.012	2.839	3.072
Bias variance				0.443	0.570	0.599	0.667	0.434	0.485
VB or PROX SE				0.782	0.739	0.637	0.543	0.330	1.031
Variance/VB or PROX SE				0.57	0.77	0.94	1.23	1.31	0.47
VF				0.834	0.826	0.785	0.771	0.732	
Variance/VF				0.53	0.69	0.76	0.86	0.59	
VP				1.009	1.054	0.998	1.029	0.816	
Bias mean	63	3.00	2.650	2.767	2.933	2.982	3.036	2.829	3.016
Bias variance			0.428	0.506	0.642	0.623	0.728	0.475	0.513
VB or PROX SE			0.846	0.742	0.708	0.603	0.541	0.312	0.989
Variance/VB or PROX SE			0.51	0.68	0.91	1.03	1.35	1.52	0.52
VF			0.882	0.794	0.787	0.766	0.771	0.731	
Variance/VF			0.49	0.64	0.81	0.81	0.94	0.65	
VP			0.994	0.929	1.012	0.973	1.046	0.817	

Table 6: Behavior of standard error variance estimates across subsets.
Simulations based on Focus group size = 25, Reference group = 175,
both groups with normal random distribution (0,1).
Unbiased items are uniformly distributed (-1,1).
Biased item with difficulty 0 (Reference), simulated bias (Focus).